

# **EDS ACTIVITY**

# **ASSIGNMENT[1]:**

**Name: Nikhil Jaiswal**

**Roll No: ME49**

**Batch: ME3**

**PRN: 202401090097**

**Subject: EDS**

**Ques:** Formulate 20 problem statements for a given dataset using Numpy and Pandas and Apply Numpy and pandas methods to find the solution for the formulated problem statements.

**Dataset:** Cricket World Cup Dataset

**Problem Statement 1:**

Load the dataset and show first 5 rows

**Code and output Screenshot:**

The screenshot shows a VS Code editor window with a file explorer on the left and a terminal at the bottom. The file explorer shows a project named 'EDS1' with a subfolder 'assignment' containing files '1.py' through '20.py' and a file 'CWC2023.csv'. The editor is open to '1.py', which contains the following Python code:

```
1 #Load the dataset and show first 5 rows.
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 print(df.head())
```

The terminal at the bottom shows the command prompt output for running the script:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/1.py
y
  Match ID  Match Date  Match Time  City  ...  Umpire 1  Umpire 2  TV Umpire  Match Referee
0         1    05-10-2023    2:00 PM  Ahmedabad  ...    Kumar Dharماسena    Nitin Menon    Paul Wilson    Javagal Srinath
1         2    06-10-2023    2:00 PM   Hyderabad  ...    Adrian Holdstock    Chris Brown    Rod Tucker    Jeff Crowe
2         3    07-10-2023    2:00 PM  Dharamashala  ...         Joel Wilson    Kumar Dharماسena    Ahsan Raza    Andy Pycroft
3         4    07-10-2023    2:00 PM    Delhi  ...    Richard Illingworth    Sharfuddoula    Michael Gough    Javagal Srinath
4         5    08-10-2023    2:00 PM    Chennai  ...    Chris Gaffaney    Richard Kettleborough    Chris Brown    Richie Richardson

[5 rows x 44 columns]
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 2:

Get dataset shape

## Code and output Screenshot:

The screenshot shows a Visual Studio Code editor window with a dark theme. The Explorer sidebar on the left displays a file tree for a project named 'EDS1', containing a folder 'assignment' with files '1.py' through '20.py', and a file 'CWC2023.csv'. The main editor area has three tabs: '1.py', '2.py' (active), and '4.py'. The active tab '2.py' contains the following Python code:

```
1 #Get dataset shape:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 print("Shape:", df.shape)
6 print("As array:", np.array(df).shape)
```

Below the code editor is a terminal window with tabs for 'PROBLEMS', 'OUTPUT', 'DEBUG CONSOLE', 'TERMINAL' (active), and 'PORTS'. The terminal shows the command to run the script and its output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/2.py
Shape: (48, 44)
As array: (48, 44)
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

The status bar at the bottom indicates the current cursor position is at Line 6, Column 40, with 4 spaces, using UTF-8 encoding and CRLF line endings. The editor is configured for Python 3.13.2.

## Problem Statement 3:

List all unique teams

## Code and output Screenshot:

The screenshot shows a Visual Studio Code editor window with a dark theme. The Explorer panel on the left shows a project named 'EDS1' with a subfolder 'assignment' containing files 1.py through 20.py and a file 'CWC2023.csv'. The file '3.py' is selected and open in the editor. The code in '3.py' is as follows:

```
1 #List all unique teams:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 teams = np.unique(pd.concat([df['Team A'], df['Team B']]))
6 print("Teams:", teams)
7
```

The bottom panel shows the 'TERMINAL' tab with the following output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & c:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/2.py
Shape: (48, 44)
As array: (48, 44)
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

The status bar at the bottom indicates the current position is Line 6, Column 23, with 4 spaces, UTF-8 encoding, CRLF line endings, Python 3.13.2, and the Go Live extension.

## Problem Statement 4:

Count matches per team

## Code and output Screenshot:

The screenshot shows a VS Code editor window with a file explorer on the left and a terminal at the bottom. The file explorer shows a project named 'EDS1' with a subfolder 'assignment' containing 20 Python files (1.py to 20.py) and a file 'CWC2023.csv'. The main editor window displays the code for '4.py'. The code is a Python script that reads a CSV file, processes the data to find unique teams, and prints the number of matches per team. The terminal at the bottom shows the command to run the script and its output, which is a dictionary mapping team names to their match counts.

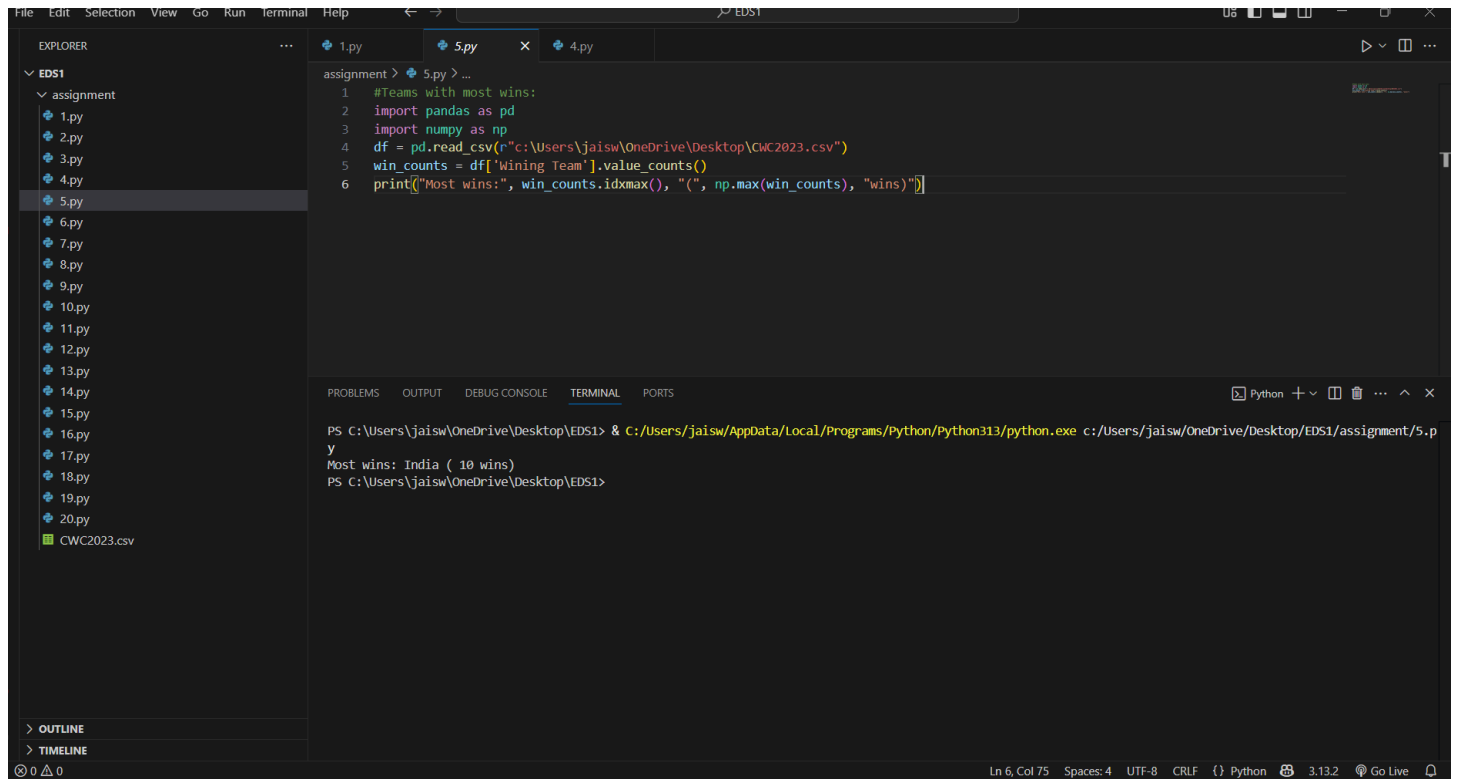
```
1 #Count matches per team
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaisw\OneDrive\Desktop\CWC2023.csv")
5 # First get all unique teams from both columns
6 all_teams = np.unique(np.concatenate([df['Team A'], df['Team B']]))
7 team_counts = np.add(
8     np.unique(df['Team A'], return_counts=True)[1],
9     np.unique(df['Team B'], return_counts=True)[1]
10 )
11
12 print("Matches per team:", dict(zip(all_teams, team_counts)))
```

```
PS C:\Users\jaisw\OneDrive\Desktop\EDS1> & C:/Users/jaisw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaisw/OneDrive/Desktop/EDS1/assignment/4.py
Matches per team: {'Afghanistan': np.int64(9), 'Australia': np.int64(11), 'Bangladesh': np.int64(9), 'England': np.int64(9), 'India': np.int64(11), 'Netherlands': np.int64(9), 'NewZealand': np.int64(10), 'Pakistan': np.int64(9), 'South Africa': np.int64(10), 'Sri Lanka': np.int64(9)}
PS C:\Users\jaisw\OneDrive\Desktop\EDS1>
```

## Problem Statement 5:

Teams with most wins

Code and output Screenshot:



The image shows a Visual Studio Code editor window with a dark theme. On the left, the Explorer sidebar shows a file tree with a folder named 'EDS1' containing an 'assignment' subfolder. The 'assignment' folder contains 20 Python files (1.py to 20.py) and a CSV file named 'CWC2023.csv'. The file '5.py' is selected and open in the main editor. The code in '5.py' is as follows:

```
1 #Teams with most wins:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 win_counts = df['Wining Team'].value_counts()
6 print("Most wins:", win_counts.idxmax(), "(", np.max(win_counts), "wins)")
```

Below the editor, the Terminal panel is active, showing the command to run the script and its output:

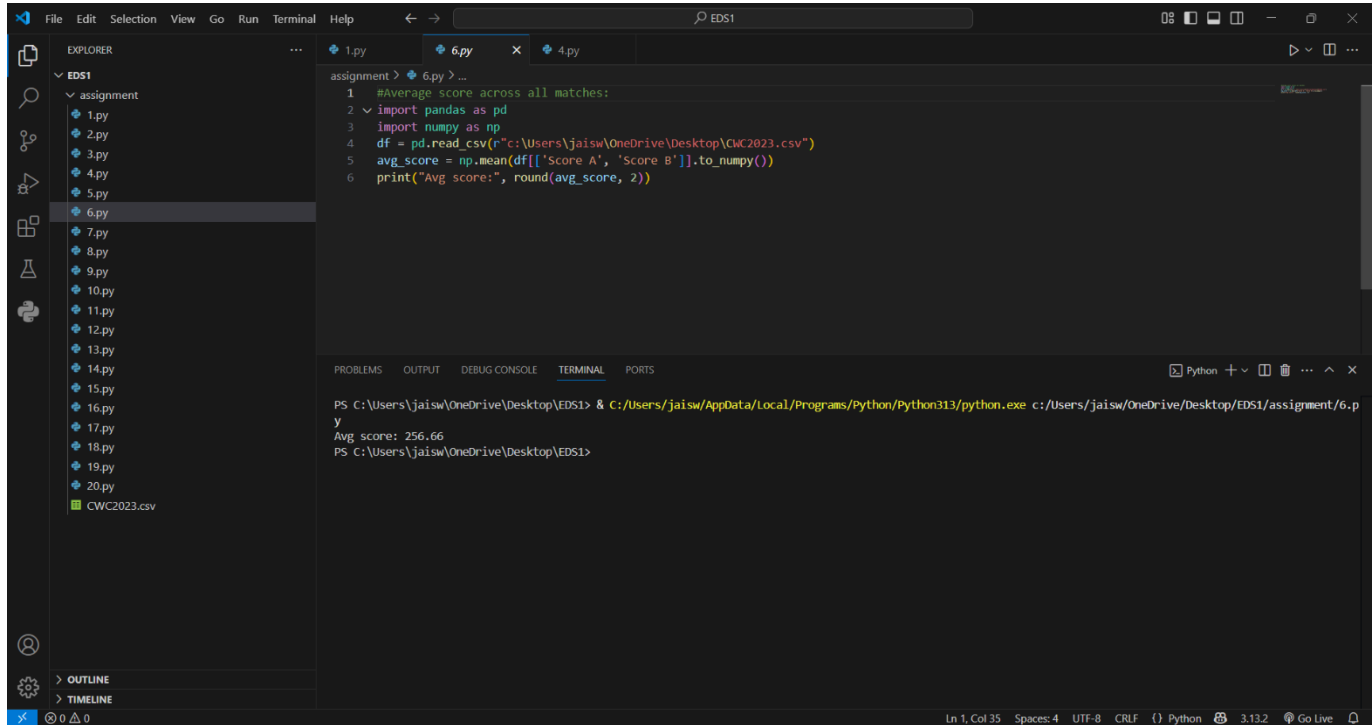
```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/5.py
Most wins: India ( 10 wins)
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

The status bar at the bottom indicates the current line and column (Ln 6, Col 75), encoding (UTF-8), line endings (CRLF), and the Python interpreter path.

## Problem Statement 6:

Average score across all matches

## Code and output Screenshot:



The screenshot shows a Visual Studio Code editor window with a file explorer on the left and a terminal at the bottom. The file explorer shows a project named 'EDS1' with a subfolder 'assignment' containing 20 Python files (1.py to 20.py) and a file 'CWC2023.csv'. The file '6.py' is selected. The editor displays the following Python code:

```
1 #Average score across all matches:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 avg_score = np.mean(df[['Score A', 'Score B']].to_numpy())
6 print("Avg score:", round(avg_score, 2))
```

The terminal at the bottom shows the command to run the script and its output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/6.py
Avg score: 256.66
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 7:

Highest total runs in match

## Code and output Screenshot:



The screenshot shows the Visual Studio Code interface with a file explorer on the left displaying a directory named 'EDS1' containing an 'assignment' folder and 20 Python files (1.py to 20.py), along with a 'CWC2023.csv' file. The main editor window shows the code for '7.py' with the following content:

```
1 #Highest total runs in match
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 df['Total Runs'] = np.add(df['Score A'], df['Score B'])
6 print("Max runs:", np.max(df['Total Runs']))
```

Below the editor, the terminal window shows the execution command and output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/7.py
Max runs: 771
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 8:

Matches with 300+ scores

Code and output Screenshot:

The screenshot shows the Visual Studio Code interface with the file explorer on the left displaying the same 'EDS1' directory structure. The main editor window shows the code for '8.py' with the following content:

```
1 #Matches with 300+ scores:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 high_scoring = df[np.any(df[['Score A', 'Score B']] > 300, axis=1)]
6 print(f"300+ matches: {len(high_scoring)}")
```

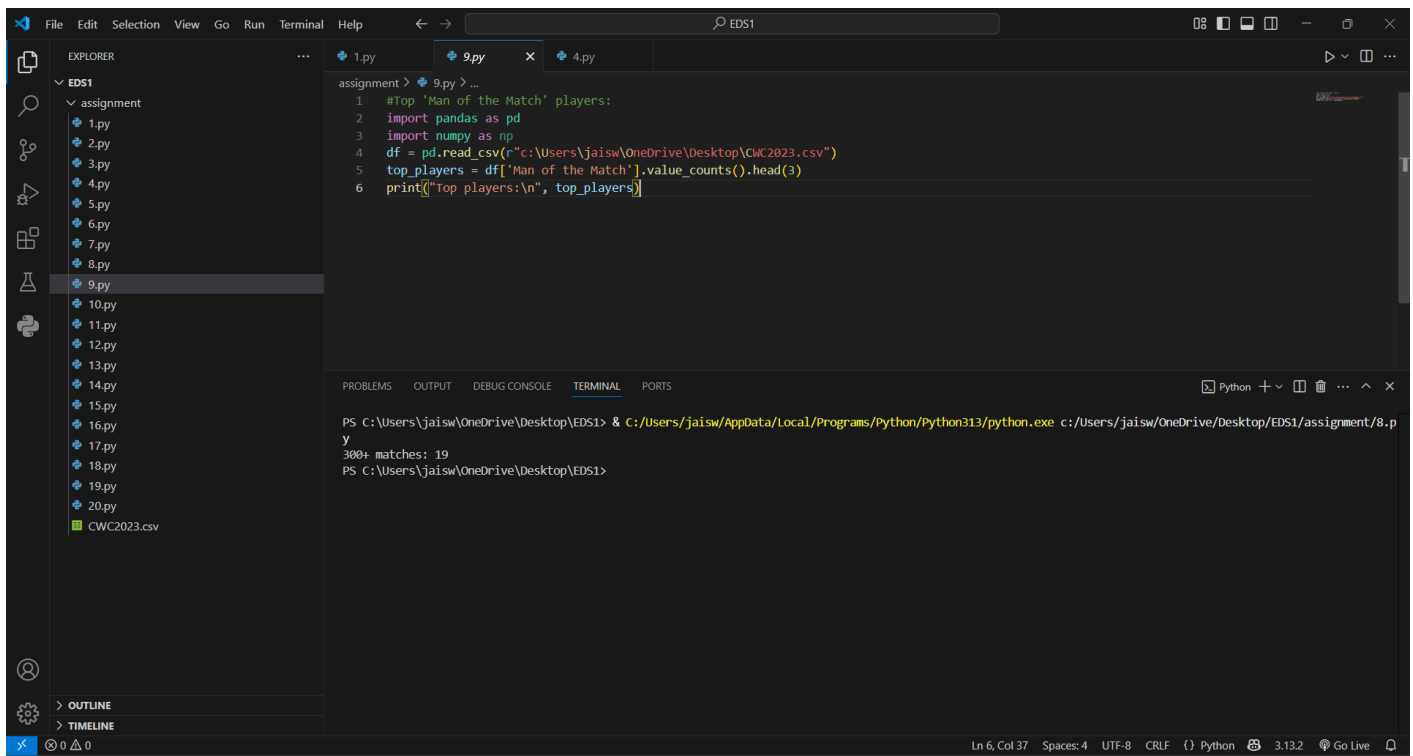
Below the editor, the terminal window shows the execution command and output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/8.py
300+ matches: 19
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 9:

Top 'Man of the Match' players

## Code and output Screenshot:



The screenshot shows a Visual Studio Code editor window with a file explorer on the left and a code editor in the center. The file explorer shows a folder named 'EDS1' containing a subfolder 'assignment' with files 1.py through 20.py and a file 'CWC2023.csv'. The code editor shows a Python script named '9.py' with the following code:

```
1 #Top 'Man of the Match' players:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 top_players = df['Man of the Match'].value_counts().head(3)
6 print("Top players:\n", top_players)
```

The terminal at the bottom shows the command to run the script and its output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/9.py
300+ matches: 19
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 10:

Count of player awards

## Code and output Screenshot:

The screenshot shows a VS Code editor with a file explorer on the left containing a folder named 'EDS1' with a subfolder 'assignment' and files 1.py through 20.py. The main editor displays '10.py' with the following code:

```
1 #Count of player awards:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 awards = np.unique(df['Man of the Match'], return_counts=True)
6 print("Player awards:", dict(zip(awards[0], awards[1])))
```

The terminal at the bottom shows the command to run the script and its output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/10.py
Player awards: {'Adam Zampa': np.int64(2), 'Aiden Markram': np.int64(1), 'Ben Stokes': np.int64(1), 'David Warner': np.int64(1), 'David Willey': np.int64(1), 'Dawid Malan': np.int64(1), 'Fakhar Zaman': np.int64(2), 'Fazalhaq Farooqi': np.int64(1), 'Glenn Maxwell': np.int64(2), 'Glenn Phillips': np.int64(1), 'Heinrich Klaasen': np.int64(1), 'Ibrahim Zadran': np.int64(1), 'Jasprit Bumrah': np.int64(1), 'KL Rahul': np.int64(1), 'Lahiru Kumara': np.int64(1), 'Lockie Ferguson': np.int64(1), 'Mehidy Hasan Miraz': np.int64(1), 'Mitchell Marsh': np.int64(1), 'Mitchell Santner': np.int64(1), 'Mohammad Nabi': np.int64(1), 'Mohammad Rizwan': np.int64(1), 'Mohammed Shami': np.int64(3), 'Mujeeb Ur Rahman': np.int64(1), 'Paul van Meekeren': np.int64(1), 'Quinton de Kock': np.int64(2), 'Rachin Ravindra': np.int64(1), 'Rassie van der Dussen': np.int64(2), 'Rohit Sharma': np.int64(2), 'Sadeera Samarawickrama': np.int64(1), 'Saud Shakeel': np.int64(1), 'Scott Edwards': np.int64(1), 'Shakib Al Hasan': np.int64(1), 'Shreyas Iyer': np.int64(1), 'Tabraiz Shamsi': np.int64(1), 'Travis Head': np.int64(3), 'Trent Boult': np.int64(1), 'Virat Kohli': np.int64(2)}
```

## Problem Statement 11:

Toss Winner Match Wins

Code and output Screenshot:

The screenshot shows a VS Code editor with a file explorer on the left containing a folder named 'EDS1' with a subfolder 'assignment' and files 1.py through 20.py. The main editor displays '11.py' with the following code:

```
1 #Toss Winner Match Wins:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 toss_wins = np.sum(df['Toss Winner'] == df['Winning Team'])
6 print("Toss winners won:", toss_wins, "/", len(df))
```

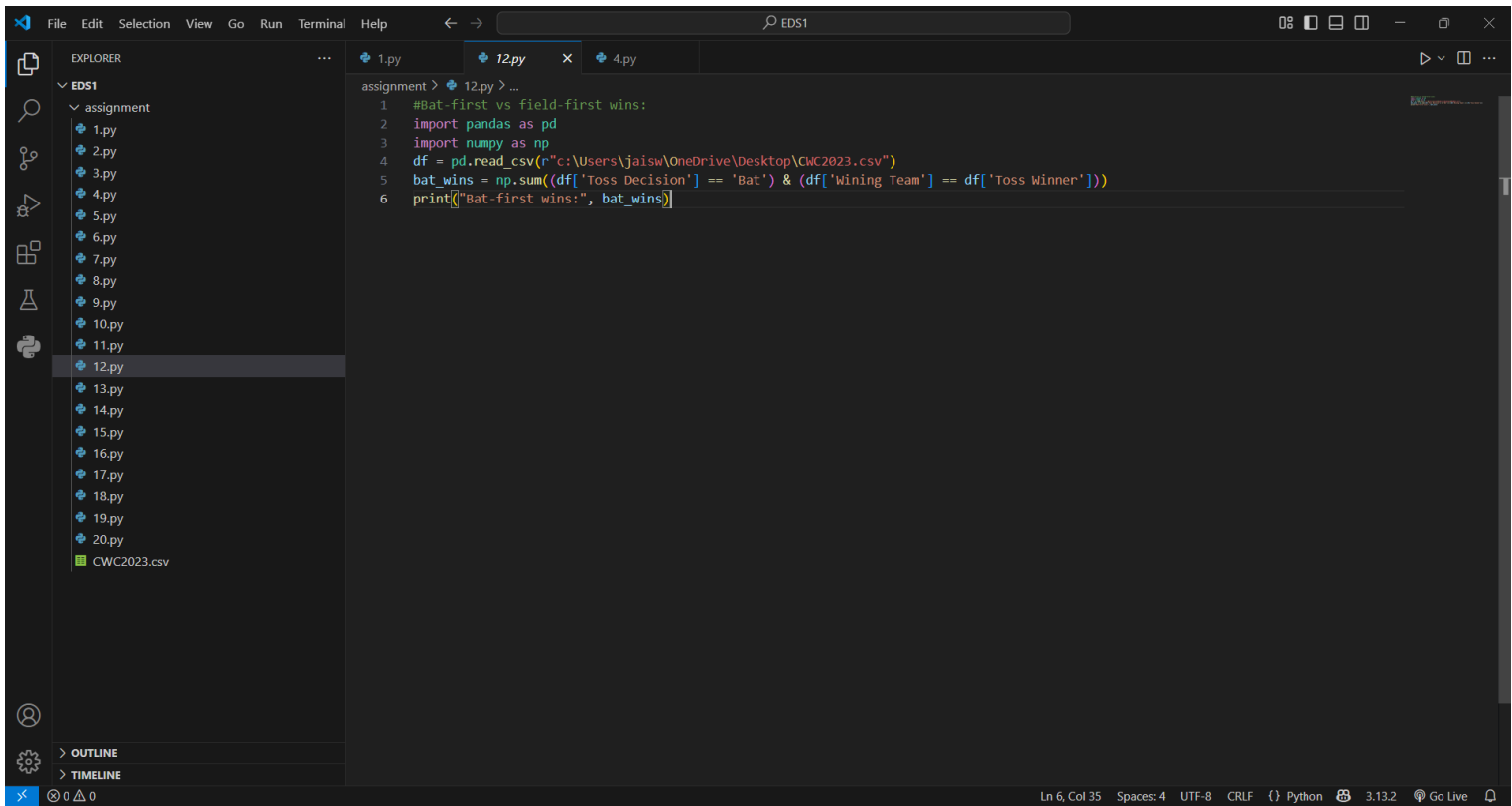
The terminal at the bottom shows the command to run the script and its output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/11.py
Toss winners won: 19 / 48
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 12:

Bat-first vs field-first wins

## Code and output Screenshot:

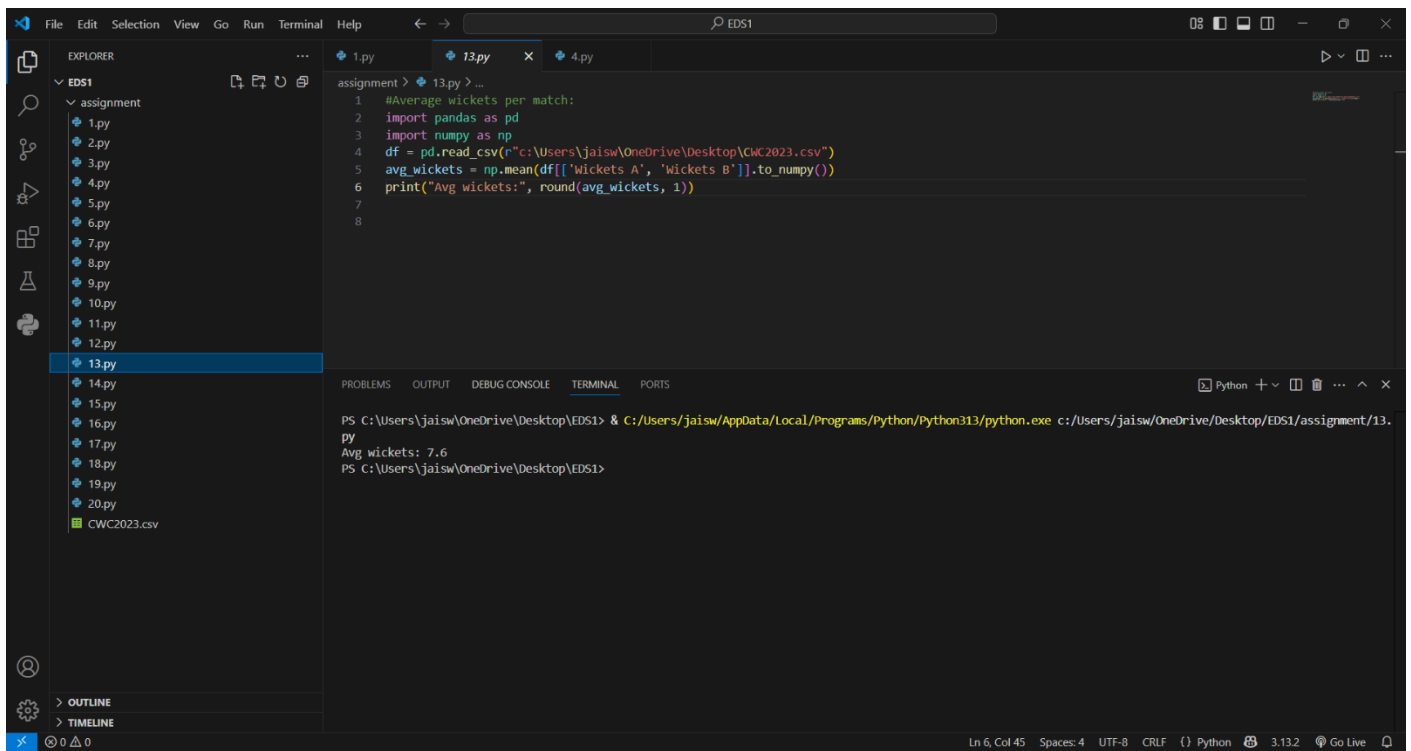


```
1 #Bat-first vs field-first wins:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 bat_wins = np.sum((df['Toss Decision'] == 'Bat') & (df['winning Team'] == df['Toss Winner']))
6 print("Bat-first wins:", bat_wins)
```

## Problem Statement 13:

Average wickets per match

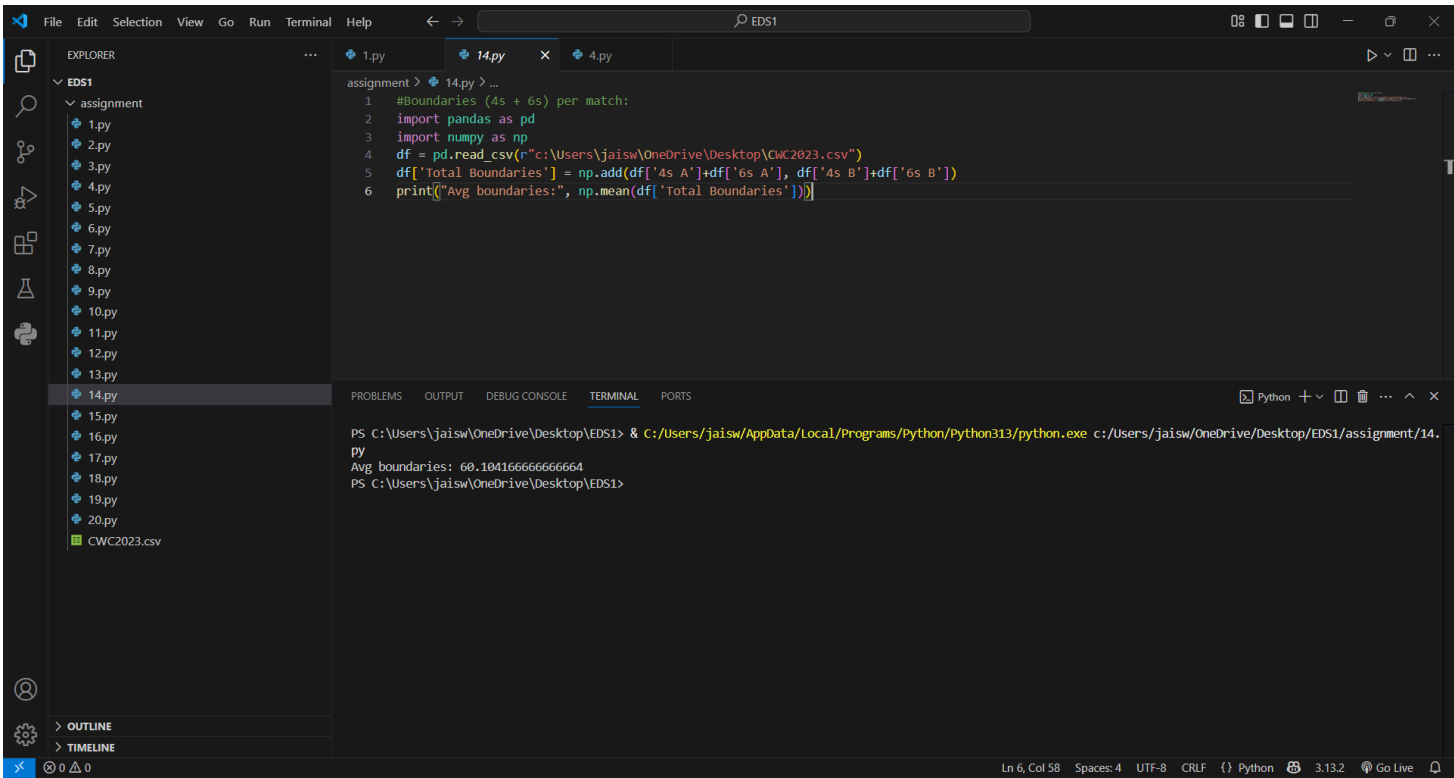
## Code and output Screenshot:



**Problem Statement 14:**

Boundaries (4s + 6s) per match

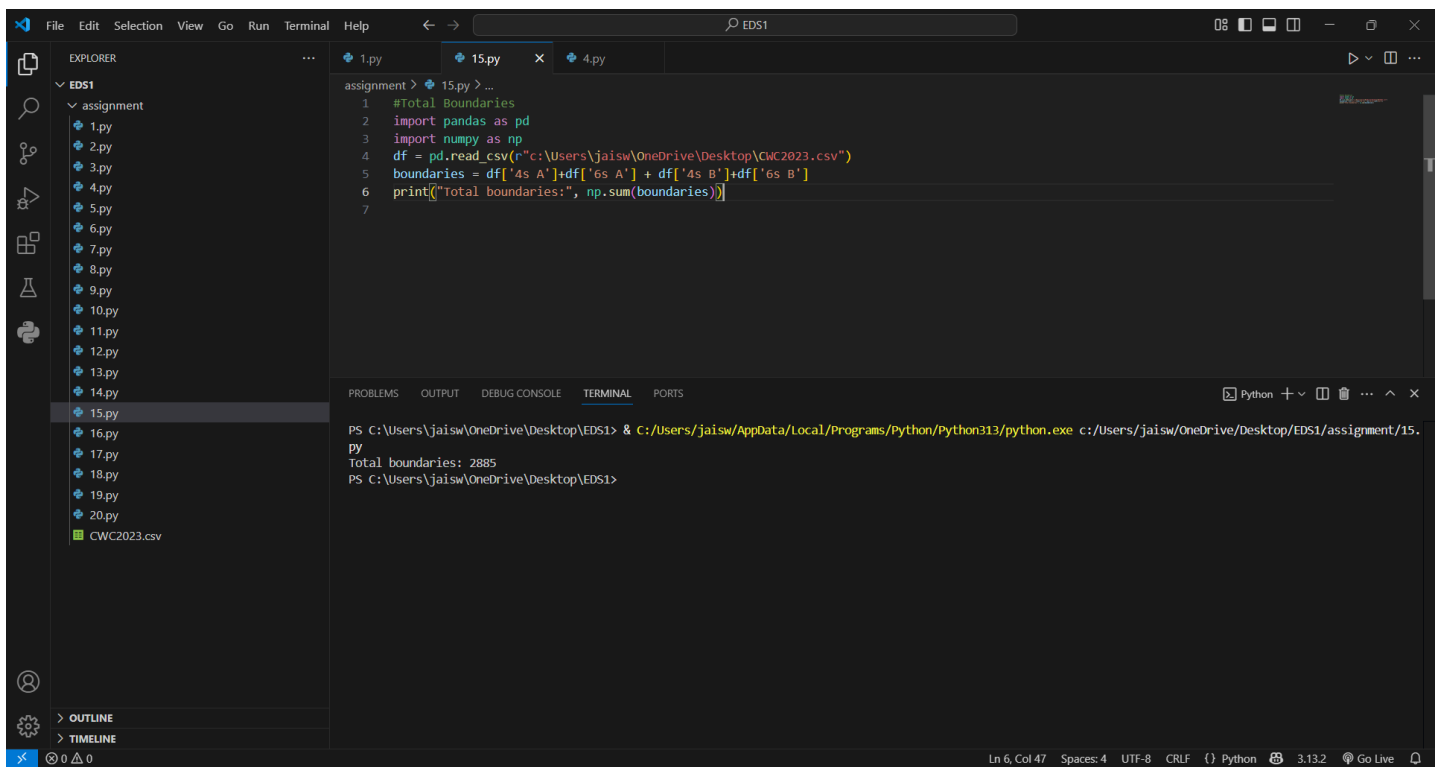
**Code and output Screenshot:**



## Problem Statement 15:

Total Boundaries

## Code and output Screenshot:



The screenshot shows a Visual Studio Code editor window with a dark theme. The Explorer panel on the left shows a file tree with a folder named 'EDS1' containing a subfolder 'assignment' with files 1.py through 20.py, and a file 'CWC2023.csv'. The file '15.py' is selected. The main editor area shows the code for '15.py':

```
1 #Total Boundaries
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 boundaries = df['4s A'] + df['6s A'] + df['4s B'] + df['6s B']
6 print("Total boundaries:", np.sum(boundaries))
7
```

Below the code editor is the TERMINAL panel, which shows the command prompt output:

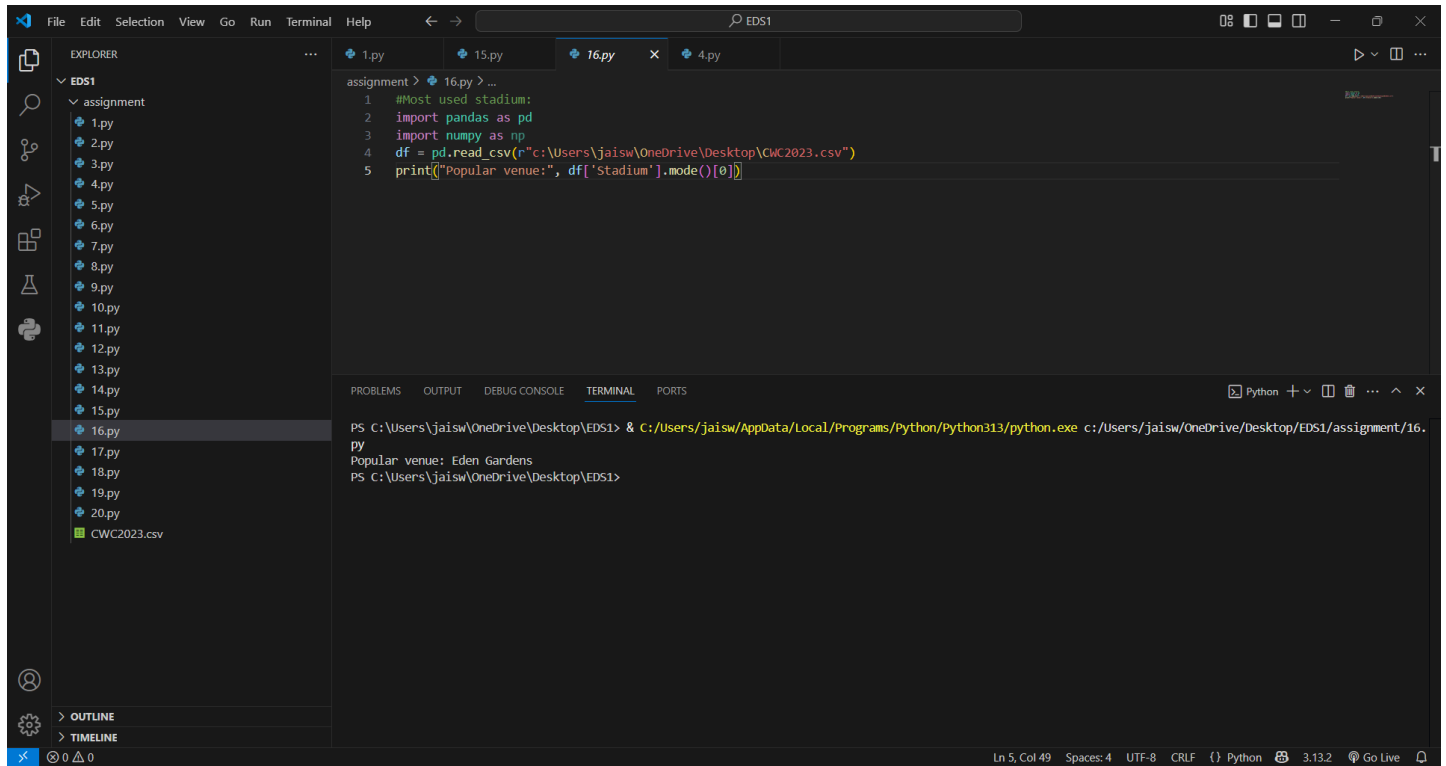
```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/15.py
Total boundaries: 2885
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

The status bar at the bottom indicates the current line and column (Ln 6, Col 47), encoding (UTF-8), line endings (CRLF), and the Python interpreter path (Python 3.13.2).

## Problem Statement 16:

Most used in stadium

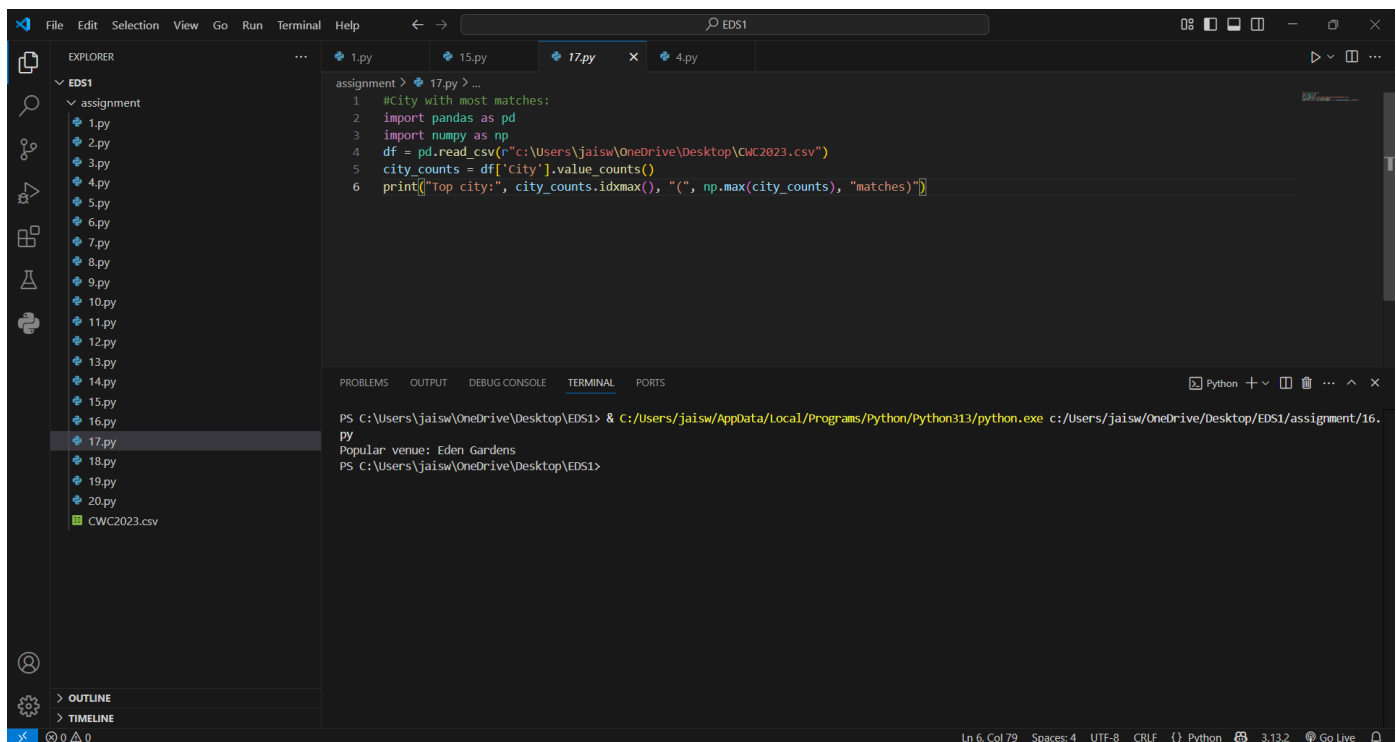
## Code and output Screenshot:



## Problem Statement 17:

City with most matches

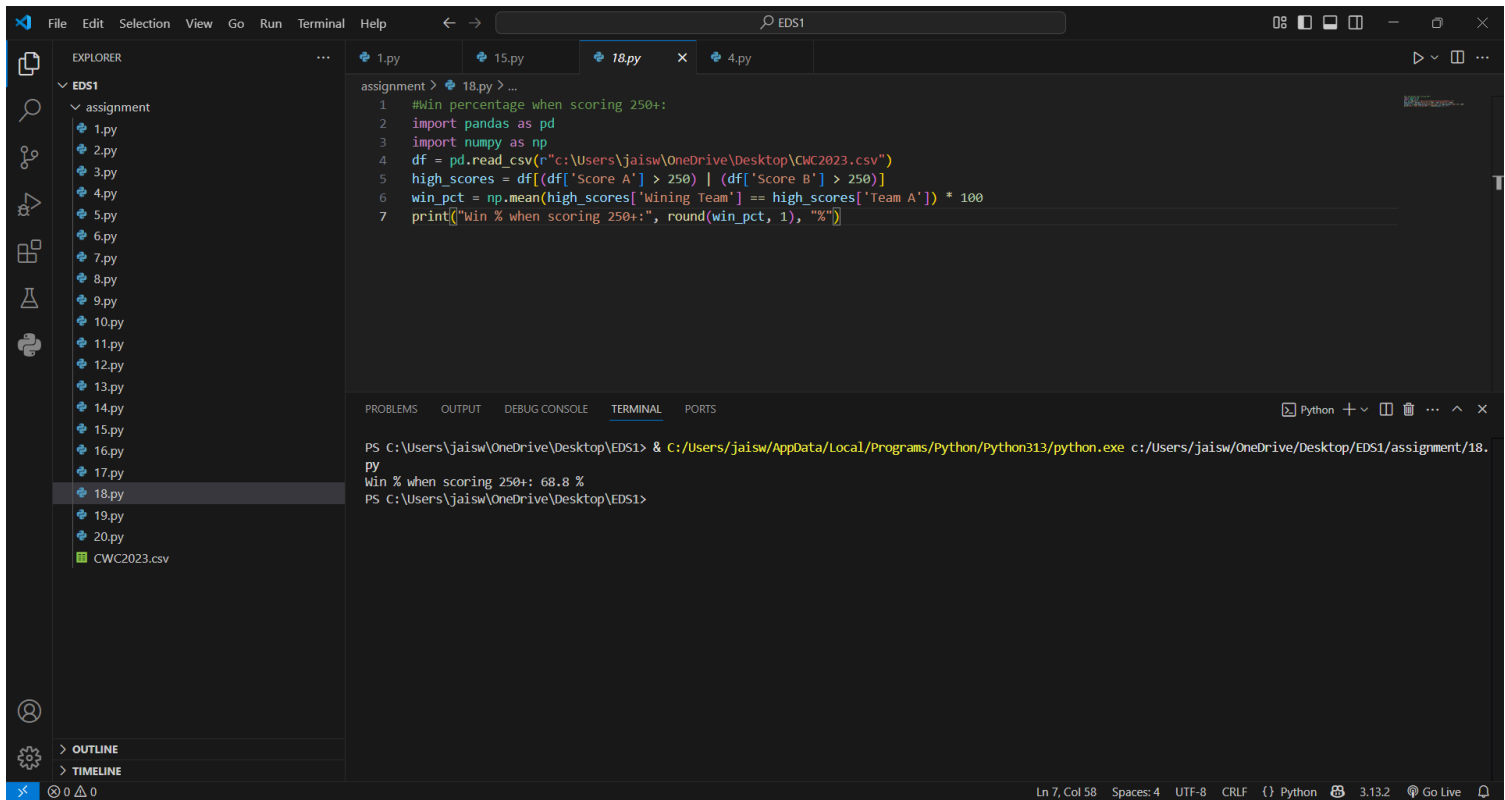
## Code and output Screenshot:



## Problem Statement 18:

Win percentage when scoring 250+

## Code and output Screenshot:



The screenshot shows a Visual Studio Code editor window with a dark theme. The Explorer panel on the left shows a project named 'EDS1' with a folder 'assignment' containing files 1.py through 20.py and a file 'CWC2023.csv'. The file '18.py' is selected and open in the editor. The code in '18.py' is as follows:

```
1 #Win percentage when scoring 250+:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"C:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 high_scores = df[(df['Score A'] > 250) | (df['Score B'] > 250)]
6 win_pct = np.mean(high_scores['Winning Team'] == high_scores['Team A']) * 100
7 print("Win % when scoring 250+:", round(win_pct, 1), "%")
```

The TERMINAL panel at the bottom shows the command prompt output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:/Users/jaishw/AppData/Local/Programs/Python/Python313/python.exe c:/Users/jaishw/OneDrive/Desktop/EDS1/assignment/18.py
Win % when scoring 250+: 68.8 %
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

The status bar at the bottom indicates the cursor is at line 7, column 58, with 4 spaces, UTF-8 encoding, CRLF line endings, and the file is a Python script.

## Problem Statement 19:

Closest match (smallest margin):

## Code and output Screenshot:



The screenshot shows the VS Code editor with a file explorer on the left. The file explorer shows a folder named 'EDS1' containing a subfolder 'assignment' and a file 'CWC2023.csv'. The 'assignment' folder contains files 1.py through 20.py. The file '19.py' is selected and open in the editor. The code in '19.py' is as follows:

```
1 #closest_match (smallest margin):
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 df['Margin'] = df['Margin'].str.extract('(\d+)').astype(float)
6 close_match = df.iloc[np.argmin(df['Margin'])]
7 print("closest match:", close_match['Team A'], "vs", close_match['Team B'])
```

The terminal at the bottom shows the command to run the script and the output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:\Users\jaishw\AppData\Local\Programs\Python\Python313\python.exe c:\Users\jaishw\OneDrive\Desktop\EDS1\assignment\19.py
c:\Users\jaishw\OneDrive\Desktop\EDS1\assignment\19.py:5: SyntaxWarning: invalid escape sequence '\d'
  df['Margin'] = df['Margin'].str.extract('(\d+)').astype(float)
Closest match: Pakistan vs South Africa
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```

## Problem Statement 20:

Extras analysis

Code and output Screenshot:

The screenshot shows the VS Code editor with a file explorer on the left. The file explorer shows a folder named 'EDS1' containing a subfolder 'assignment' and a file 'CWC2023.csv'. The 'assignment' folder contains files 1.py through 20.py. The file '20.py' is selected and open in the editor. The code in '20.py' is as follows:

```
1 #Extras analysis:
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv(r"c:\Users\jaishw\OneDrive\Desktop\CWC2023.csv")
5 total_extras = np.sum(df['Extras A'] + df['Extras B'])
6 print("Total extras in tournament:", total_extras)
```

The terminal at the bottom shows the command to run the script and the output:

```
PS C:\Users\jaishw\OneDrive\Desktop\EDS1> & C:\Users\jaishw\AppData\Local\Programs\Python\Python313\python.exe c:\Users\jaishw\OneDrive\Desktop\EDS1\assignment\20.py
Total extras in tournament: 1215
PS C:\Users\jaishw\OneDrive\Desktop\EDS1>
```